Weekly Report

May 20th, 2018

Done:

1. Finish the coursera course on "Object detection and localization", implement the state-of-art YOLO algorithm with keras.
2. Pull down the from DB, match with crawled data. Clean the erroneous records. Two main issues:
    a. Same id, different names (typos or different complex of fire)
    b. Missing records of DB data or crawled data. We must manually check all these issues and correct them. Luckily the number of issues is <100.
3. Listen to a talk about "Towards Flexible but controllable language generation". It's more about the heuristics behind the construction of their NLP Neural Network.
4. Had a very brief talk with Prof. Ma during lunch about the job I've done. Since I'm doing some summarization of the data, he said I can try visualization telling, and do some layout algorithm stuff. I had a talk to the storytelling expert Chris Bryan in this lab, centered on his InfoVis16 storytelling paper, and learn several potential topics:
   (a) interesting/anomaly events detection and annotation. Find these POIs (based on several available implementation like uncertainty and KDE, and probably more algorithms on these geodata like hotspot formation), generate annotations, label them on the appropriate place (for our data, on the geospatial map, or on the infovis summarization visualization)?
   (b) data comics/graph comics. One good example is CHI16 paper: Telling Stories about Dynamic Networks with Graph Comics. This work is all done manually. In Chris's paper, he used snapshots to realize a similar functionality. Many timestamp-selection methods are provided. For our data, we have explored some snapshots display vis about the fire area uncertainty results.
   (c) annotation filtering spatially. In his paper, there is an algorithm help users to filter annotations based on the rank, the spatial density information. Our data has inherent spatial information, which can be or combined with other annotations to do such a filtering.
   (d) summarize annotations (condense them into a story). This may require NLP techniques.
   (e) dynamically generate and remove annotations while scrolling the timeline. This should be a required functionality in the system.
   A more detailed discussion will be on next week.
5. The graph exploration topic: Tarik mentioned that the user study based on user interaction history is not easy to conduct. The reason is that we can't get a ground truth of recommendation on such interaction history.
6. Finish revising the Chap.12 of visualization book.

To Do:

1. Talk to Prof. Ma about the ideas above, try to narrow down the topics.

2. Clean the data (still on the California fire data). Correlate them with
3. All the backend switch to python Falcon.

Paper reading:

1. 大黑书最后补充了 4 篇图不确定性可视化文章细节.

2. 上面提到的 storytelling infovis16 的文章，一些重要思想已经在上面提到了.

3. VISAGE: Interactive Visual Graph Querying: 类似于交互式构建一个 onto 来做查询，不需要再写复杂的查询语句

4. Apolo: Making Sense of Large Network Data(Making Sense of Large Network Data: Combining Rich User Interaction and Machine Learning). 其方法有别于 Tarik 的文章在于，Tarik 文章对于数据添加额外信息并支持进一步分析很少，而且还比较关注 layout；本文不关注 layout，但是在增加信息、增量推荐上做了比较多的花样. 本文方法需要用户对数据熟悉、并通过构建 exemplar 来推荐更多；在此基础上也要做合理的分类.